AFHRL-TP-83-64

(13)

AIR FORCE

HUMAN RESOURCES

COGNITIVE ORGANIZATION AS A FUNCTION OF FLYING EXPERIENCE

By

Roger W. Schvaneveldt
Timothy J. Breen
Nancy M. Cooke

Department of Psychology
New Mexico State University
Las Cruces, New Mexico 88003

Francis T. Durso

Department of Psychology
University of Oklahoma
Norman, Oklahoma 73019

Timothy E. Goldsmith

Department of Psychology
University of New Mexico
Albuquerque, New Mexico 87131

Richard G. Tucker

479th TTW/Training Analysis
Holloman Air Force Base, New Mexico 88330

Joseph C. DeMaio

OPERATIONS TRAINING DIVISION
Williams Air Force Base, Arizona 85224

May 1984

Final Technical Paper

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this paper, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This paper has been reviewed and is approved for publication.

JOSEPH C. DeMAIO
Contract Monitor

MILTON E. WOOD, Technical Director
Operations Training Division

CARL D. ELIASON, Colonel, USAF
Chief, Operations Training Division

AIR FORCE HUMAN RESOURCES LABORATORY
Brooks Air Force Base, Texas 78235

ERRATUM

Schvaneveldt, R.W., Breen, T.J., Cooke, N.M., Durso, F.T., Goldsmith, T.E., Tucker, R.G., & DeMaio, J.C. Cognitive organization as a function of flying experience. AFHRL-TP-83-64. Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory, May 1984.

On DD Form 1473, block 5, change AFHRL-TR-83-64 to read AFHRL-TP-83-64.

LOU ELLIOTT
Chief, Technical Editing Office

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE:

3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S):

5. MONITORING ORGANIZATION REPORT NUMBER(S): AFHRL-TP-83-64

6a. NAME OF PERFORMING ORGANIZATION: Department of Psychology

6b. OFFICE SYMBOL *(If applicable)*:

7a. NAME OF MONITORING ORGANIZATION: Operations Training Division

6c. ADDRESS *(City, State and ZIP Code)*:
New Mexico State University
Las Cruces, New Mexico 88003

7b. ADDRESS *(City, State and ZIP Code)*:
Air Force Human Resources Laboratory
Williams Air Force Base, Arizona 85224

8a. NAME OF FUNDING/SPONSORING ORGANIZATION: HQ Air Force Human Resources Laboratory (AFSC)

8b. OFFICE SYMBOL *(If applicable)*:

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F33615-80-C-0004

8c. ADDRESS *(City, State and ZIP Code)*:
Brooks Air Force Base, Texas 78235

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2313	T3	13

11. TITLE *(Include Security Classification)*: Cognitive Organization as a Function of Flying Experience

12. PERSONAL AUTHOR(S): Roger W. Schvaneveldt, Nancy M. Cooke, Timothy E. Goldsmith, Joseph C. DeMaio, Timothy J. Breen, Francis T. Durso, Richard G. Tucker

13a. TYPE OF REPORT: Final

13b. TIME COVERED: FROM ________ TO ________

14. DATE OF REPORT *(Yr., Mo., Day)*: May 1984

15. PAGE COUNT: 54

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB. GR.
0508		
0509		

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*:
critical flight information, flying training, general weighted networks, multidimensional scaling, pilot expertise, psychometric scaling, training design

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

This report reviews work in defining and measuring conceptual structures of critical flight information in Air Force fighter pilots. Groups of pilots with widely varying expertise were tested. Cognitive structures were defined by multidimensional scaling (MDS) and general weighted networks (GWN). The structures were validated by recovering the experience differences among the pilots from their conceptual structures. Group membership can be predicted from a person's conceptual structure. The techniques employed permit detailed analyses of individual differences, and they point to factors distinguishing expert and novice pilots. The GWN analysis led to the identification of specific points of agreement and disagreement in the conceptual organization of novice and expert pilots. Pilots do have measurable cognitive structures for organizing flight-related information. These structures are measurably different for individuals with differing flight experience. The techniques used here produce descriptions of conceptual structure that may have application in training and assessing individual differences in the development of expert conceptual structures.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED [X] SAME AS RPT. [ ] DTIC USERS [ ]

21. ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL: Nancy A. Perrigo

22b. TELEPHONE NUMBER *(Including Area Code)*: 512-536-3877

22c. OFFICE SYMBOL: AFHRL/TSR

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

AFHRL Technical Paper 83-64 May 1984

COGNITIVE ORGANIZATION AS A FUNCTION OF FLYING EXPERIENCE

By

Roger W. Schvaneveldt
Timothy J. Breen
Nancy M. Cooke

Department of Psychology
New Mexico State University
Las Cruces, New Mexico 88003

Francis T. Durso

Department of Psychology
University of Oklahoma
Norman, Oklahoma 73019

Timothy E. Goldsmith

Department of Psychology
University of New Mexico
Albuquerque, New Mexico 87131

Richard G. Tucker

479th TTW/Training Analysis
Holloman Air Force Base, New Mexico 88330

Joseph C. DeMaio

OPERATIONS TRAINING DIVISION
Williams Air Force Base, Arizona 85224

Reviewed and submitted for publication by

Robert B. Bunker
Chief, Technology Development Branch

This publication is primarily a working paper.
It is published solely to document work performed.

PREFACE

The central goal of this effort has been to measure the structure of flight-related concepts for Air Force fighter pilots. The project has led to the development of a new method of scaling conceptual structures (general weighted networks) and has involved extensive use and novel applications of another well-established method (multidimensional scaling). This report reviews work on the measurement of conceptual structure and on assessing the reliability, validity, and applicability of particular structural descriptions. This work is part of the Air Force Human Resources Laboratory's 6.1 basic research program and is intended to advance understanding of basic cognitive dimensions in flying behavior.

Dr. Don Dearholt, Dr. Jim McDonald, Wayne Whitmore, John Yorshak, Ted Dunning, JoAnne Barnes, Debora Stefan, Yvonne Boudreau, Karen Preuss, Ann Schvaneveldt, and Melvin Tempel at New Mexico State University have made numerous contributions to this work. The cooperation of pilots and other Air Force personnel at Holloman AFB and Williams AFB and of fighter pilots of the 120th TAC Fighter Squadron at Buckley ANGB is gratefully acknowledged.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

COGNITIVE ORGANIZATION AS A FUNCTION OF FLYING EXPERIENCE

I. INTRODUCTION

In the past decade, cognitive psychologists have generated a considerable body of theory and data concerning the organization and retrieval of knowledge in human memory. Researchers in this area (which has come to be known as semantic memory) have begun to gain an understanding of the representation of knowledge by investigating the exceptionally rich data bases of natural language and natural categories. This research has demonstrated that the organization of memory exerts important influences on the encoding and retrieval of information.

Much of the research in semantic memory has focused on the influence of semantic relatedness on the speed and accuracy with which task relevant information can be retrieved from memory. Various terms, such as semantic similarity, semantic relatedness, and semantic distance, have been used to refer to the degree to which concepts are related in meaning. The distance metaphor comes from an analogy to a multidimensional space where concepts are located according to values on various dimensions of meaning. Presumably, concepts near one another in multidimensional space are more closely related to one another than are concepts that are farther apart in the space.

There have been several proposals concerning memory structures, but each one makes use of the idea that concepts in memory differ in their relatedness or psychological proximity. In network models, concepts are represented as nodes linked by labeled relations. Two concepts that are directly linked are viewed as more similar than are two concepts that are not linked or are indirectly linked (Collins & Quillian, 1969; Quillian, 1969). Similarly, in other network models, two concepts that share a number of links are viewed as more related than are two concepts that share fewer links (Collins & Loftus, 1975). In featural models (Rips, Shoben, & Smith, 1973), where concepts are represented by vectors of features, two concepts that share a number of features are viewed as more similar than are two concepts that share few, if any, features. Both network and feature theories rely on psychological proximity to predict performance in a variety of tasks.

How theorists have determined the proximity of a particular set of concepts has varied widely. Most models applied to particular domains have relied solely, or primarily, on the intuitions of the theorists. There are, however, a number of notable exceptions. Smith, Shoben, and Rips (1974) employed multidimensional scaling (MDS) procedures for a set of animal names to reveal important structural information. Similarly, Shepard (1963) and Kruskal (1977) have investigated the applicability of multidimensional spatial representations for a number of conceptual domains with some

encouraging results. Until recently, theorists assuming a general network as an underlying model were limited to intuitions or to restrictive clustering schemes that allow only hierarchical relations between concepts. Recently developed techniques, however, allow researchers to derive networks from the same proximity data employed by MDS (Durso, Schvaneveldt, & Goldsmith, 1983; Hutchinson, 1981; Schvaneveldt & Durso, 1981). The present report makes extensive use of both MDS and network techniques.

In this research, the techniques applied are those used by cognitive psychologists to define and measure conceptual structures of individuals in a restricted domain. This project investigated the conceptual structures of Air Force fighter pilots for combat situations. The central goal was to demonstrate the existence and utility of a systematic structure of flight-related concepts in the memory systems of fighter pilots. Meeting this goal required applying structural analyses to data and developing methods for assessing the validity of the structural representations. This domain is one that, in principle, has a large, rich cognitive component that should reflect important facets of conceptual structure in general. The domain is relatively self-contained and not merely an arbitrary subset of natural language. The domain allows identification of individuals that vary in their mastery of the domain and, presumably, in the nature of their conceptual structures. These variations in expertise provide one approach to validating measures of conceptual structure. Presumably, the conceptual structures of experts should differ systematically from the conceptual structure of novices.

Techniques from cognitive psychology have been employed in order to compare memory structures of experts and novices. Expert-novice studies are concerned not only with expert-novice differences in memory structure but also with the development of this structure as a novice gains skill or experience and approaches the expert level. This latter issue has important applications in training and education. Other issues in this area focus on how experts organize information in memory, expert-novice differences in performance on recall and perceptual tasks, and methods for measuring, representing, and validating structures of memory. Expert and novice studies have been conducted in domains such as chess, bridge, Go, physics, and computer programming (Adelson, 1981; Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; Engle & Bukstel, 1978; McKeithen, Reitman, Rueter & Hirtle, 1981; Reitman, 1976).

This expert-novice research has resulted in some fairly general conclusions in regard to memory and expertise. By definition, the performance of experts on the actual task in which they excel is superior to the performance of novices. Experts, however, show superior performance on recall tasks in which meaningful material is used, but are no better than novices when asked to recall the same material in a random arrangement. For instance, chess masters are

able to recall positions of pieces on a chess board much more readily if the pieces are arranged as they would be in a game situation, rather than a random arrangement. Yet, this meaningful arrangement of the pieces does not aid chess novices in recalling positions. Furthermore, as experience increases, there tends to be a greater degree of intragroup agreement, in relation to memory structure and organization. Other common findings include a larger chunk size and more chunks for the experts as compared to the novices. Chunks are units of information in which the items within a chunk are related to each other in a meaningful fashion.

Various explanations have been offered for these findings. Typically, it has been suggested that the expert is able to perceive a more global picture and, therefore, is able to encode or chunk items into larger units than is the novice. The novice has a memory structure that is not as highly organized as that of the expert and, therefore, is not able to encode as quickly or in as large units. It has also been suggested that experts have a memory structure that is hierarchically organized and, therefore, can recall a larger number of chunks. Thus, a high-level chunk may consist of a set of lower-level chunks, each containing additional chunks at a more detailed level. This type of organization can benefit experts by constantly cueing them about future chunks.

Research involving fighter personnel has traditionally dealt with perceptual and motor components of flying high speed tactical aircraft. Research taking a cognitive perspective has been rare, despite the fact that tactical flight requires a complex knowledge base from which to operate. The organization of information in memory is believed to have a critical impact on flying performance. Understanding how critical information is organized in memory can be extremely useful to training program designers and evaluators as well as to instructors and others interested in increasing the effectiveness of the pilot-aircraft system. Knowledge of how individuals develop systems for organizing critical information can be used to tailor training systems to provide students the conceptual framework that will lead to optimal learning. It may also provide a useful selection tool by allowing instructors to determine which individuals have mastered the prerequisite concepts for success in a particular training program.

II. GENERAL METHODOLOGY

Subjects

Three populations of fighter pilots were sampled for these studies. Ten Instructor Pilots (IPs) stationed at Holloman Air Force Base (AFB) and nine Air National Guard Pilots (GPs) from Buckley Air National Guard Base served as the two groups of expert pilots. The IPs averaged 2583 hours flying time and served as instructors for lead-in fighter training. The GPs averaged 6064 hours flying time but were not classroom instructors. Although the IPs and GPs were experienced pilots, their experience differed in that the GPs had little instructor experience, whereas the IPs had relatively less operational experience. The third sample consisted of 17 Undergraduate Pilot Trainees (UPs) stationed at Williams AFB. The UPs averaged 200 hours of flying time and had recently completed Undergraduate Pilot Training. Undergraduate Pilot Training precedes advanced training with specialty aircraft, thus none of the UPs had undergone fighter lead-in training. This choice of subjects seemed to be appropriate because they were expected to exhibit some, but certainly not all, of the features characteristic of expert fighter pilots. In particular, the UPs should have a good command of general flying procedures (e.g., formation flying) but little or no command of air-to-air or air-to-ground combat situations.

Materials

The development of the stimulus materials began with a task analysis of tactical flight maneuvers (Meyer, Laveson, Pape, & Edwards, 1978). According to Meyer et al. (1978) the use of a scenario was effective with tactical aircrews in establishing rapport and isolating parameters for subsequent data collection. Based on the Meyer report and through dynamic interaction with four pilots from the 449TTW, two scenarios were developed. One scenario concerned split-plane maneuvers in air-to-air combat, and the other scenario concerned the low-angle strafe maneuver in air-to-ground combat. These two scenarios were chosen, in part, because they differ greatly in inherent complexity. The split-plane scenario is inherently complex, involving several possible configurations of aircraft, instruments, and possible actions. The strafe scenario is inherently simple, involving a relatively rigid configuration of a single aircraft, instruments, and possible actions. In addition, these scenarios contain some concepts that should be well understood by the UPs and some concepts that should be relatively foreign to the UPs, who have not received fighter lead-in training.

Each scenario consisted of a set of assumptions and 30 basic concepts. The two scenarios appear in Table 1. The basic concepts served as the critical stimuli for the experiments. Thus, 30 concepts important to air-to-air combat and 30 concepts important to air-to-ground combat were examined in these studies.

Table 1. Two Scenarios with Assumptions and Basic Terms

a. Split-Plane Maneuvers

Assumptions

OFFENSIVE	AGGRESSIVE	TALLY HO	SINGLE BANDIT
KILL	COMMIT	ENGAGED	DEFENSIVE TURN
SIMILAR AIRCRAFT		IR MISSLE PARAMETERS	

Basic Concepts

LOW YO YO	HIGH YO YO	QUARTER PLANE	ASPECT ANGLE
LAG ROLL	BARREL ROLL	OVERTAKE	PURE PURSUIT
GUNS	AIRSPEED	CORNER VELOCITY	LEAD PURSUIT
G LOADING	CUTOFF	RELATIVE ENERGY	LAG PURSUIT
6 O'CLOCK	SMASH	POWER SETTING	LIFT VECTOR
SWITCHOLOGY	RADIAL G	ACCELERATION	3-9 LINE
HEAT	SNAPSHOT	VERTICAL MANEUVERING	
EXTENSION	ANGLE OFF	WEAPONS PARAMETERS	

b. Strafe Maneuver

Assumptions

CONTROLLED RANGE	PANEL/TARGET	CLEARED
TARGET ACQUISITION	SWITCHOLOGY	

Basic Concepts

BULLET IMPACT	AIM OFF POINT	DIVE ANGLE	DRIFT
GLIDEPATH	FOUL LINE	CLOSURE	GUNS
AIRSPEED	RUN-IN LINE	ALTITUDE	AIM POINT
BANK	PIPPER FIXATION	WALKING	RANGE
TRIGGER	TRACKING	PIPPER PLACEMENT	PULL-UP
RICOCHET	YAW	FINAL	FIRE
BURST	RECOVERY	BUNT	
STABILIZE	TRIM	FOUL	

Procedure

Most scaling procedures for producing structural descriptions of a set of concepts require some measure of psychological distance between the concepts. Although two general methods have been used in the literature, inter-item distance in recall protocols (e.g., Adelson, 1981) and direct judgments of pairwise similarity (e.g., Rips, et al., 1973), recent work by Cooke (1983) suggests that the latter measure provides a more sensitive and valid database for the delineation of conceptual structures. In the present project, measures of proximity were based on pairwise similarity judgments. All data were collected using a TERAK microprocessor system.

Similarity ratings. The subjects were told about similarity or relatedness ratings and the mechanical details of entering ratings on the TERAK. The scenario was then described to provide a context for rating the basic terms, and the complete set of terms to be rated was shown to allow subjects to establish some criteria for rating the pairs of concepts.

The rating task itself consisted of presenting all possible pairs of the 30 basic concepts. Subjects rated the similarity of 435 pairs of terms (i.e., 30 taken two at a time) during the session. For each pair of terms, the TERAK displayed the pair of terms to be rated, a rating scale with the numbers 0 through 9, and a bar marker to indicate the rating. Subjects were instructed that a number of factors might enter into a decision about similarity, including relatedness, co-occurrence, dependency, and contingency. They were told that the purpose was to obtain their general impressions of the similarity of two items and that they should not ponder their judgments. Subjects entered their rating by pressing a number key on the TERAK keyboard. The bar marker in the display was moved to the position corresponding to the number entered by the subject to indicate the rating given. The subject could change the rating by pressing another number key. When the subject was satisfied with the rating, pressing the space bar on the keyboard changed the display to show the next pair of items and reset the marker to the bottom of the scale. This procedure was followed until all 435 pairs had been presented. The order of the pairs was randomized for each subject, and the position of the two items in a pair was counterbalanced across subjects. A rating session required from 30 to 45 minutes to complete.

Familiarity ratings. In addition to the rating task, the UP subjects were asked to rate their familiarity with each of the concepts. This was to help determine in a rough way the level of experience with these critical concepts. UPs rated each concept on a scale of 1 to 3, where 1 indicated no familiarity, 2 indicated familiarity, and 3 indicated the concept had been used in flying.

III. PRELIMINARY DATA ANALYIS

Seven of the 10 IPs, each of the GPs, and each of the UPs supplied data in the split-plane scenario. For the strafe scenario, data were collected from 6 of the IPs and 16 of the UPs; no data were obtained from the GPs for the strafe scenario.

The obtained similarity measures were transformed into measures of psychological distance by subtracting the ratings from the maximum possible rating. The resulting numbers reflect distance, with the larger numbers representing greater psychological distance between concepts. For each scenario, for each subject, the data were placed in a 30 by 30 symmetrical matrix where all entries, other than the diagonal, represented the empirical judgment for a pair of concepts. Similar matrices of means were computed for each scenario and each group of subjects.

Correlations

A preliminary question was intended to determine the extent to which individuals agreed on the ratings obtained for each scenario. Reliable correlations were expected within each group of experts; also of interest was the extent to which students tended to agree with other students. In addition to these intra-group correlations, another item of interest was the degree of agreement between pilots from different groups. If the rating data are valid, they should reveal higher intra-group correlations compared to intergroup correlations; further, the intercorrelations between the experts, IPs and GPs, should be somewhat higher than the intercorrelations between UPs and either IPs or GPs.

The rating matrix for each subject was correlated with that of every other subject. Mean correlations appear in Table 2. Despite the magnitude of the correlations, all were statistically reliable. As shown in subsequent sections of this report, the techniques employed are quite successful even when the correlations average approximately .3. It will be demonstrated that there is sufficient information captured in the rating data to discriminate among subject populations. In addition, because these are mean correlations between individuals, the contribution of noise is maximized both within and between groups. Scaling procedures are often applied to consensus data with intersubject correlations of this magnitude or smaller. The scaling techniques employed should prove quite effective in extracting the latent structure from these data.

Table 2. Intercorrelation Matrices for Rating Data
(Entries are mean correlations)

a. Split-Plane Maneuvers

	IPs	GPs	UPs	Mean
IPs	.42	.35	.20	.32
GPs	.35	.36	.24	.32
UPs	.20	.24	.31	.25

b. Strafe Maneuver

	IPs	UPs	Mean
IPs	.44	.20	.32
UPs	.20	.32	.26

Inspection of Table 2 suggests that agreement within a group (diagonal) was consistently higher than any intercorrelation involving that group. The intracorrelations for an expert group were higher than the intercorrelations involving UPs: for IPs $z = 2.19$, $p < .05$ and $z = 4.43$, $p < .05$, in the split and strafe scenarios, respectively; for GPs $z = 2.20$, $p < .05$. However, experts tended to agree as much with experts from the other group as they did with members of their own group. The greatest consistency with members of their own group was shown by the IPs; although this was not reliably higher than the intragroup GP correlation, it was higher than the intragroup correlation for UPs, $z = 1.86$, $p < .1$ and $z = 2.06$, $p < .05$, for split and strafe, respectively. This is perhaps not surprising since instructors not only know about the execution of the maneuvers, but they also are required to organize what they know so they can communicate it to their students. Even students showed higher correlations with other students than they did with either group of experts, and this was statistically the case in comparisons with IPs: $z = 1.907$, $p < .1$ and $z = 2.10$, $p < .05$, for split and strafe, respectively. Apparently, whatever it is that students think about the concepts being rated, they share the same knowledge to some extent; however, the shared knowledge is less than that of the experts, especially the IPs.

The second point to note is that the least experienced pilots (i.e., UPs) also show lower correlations with the other groups. Table 2 shows quite clearly that IPs and GPs showed higher agreement than either UPs and IPs, $z = 2.66$ $p < .05$, or UPs and GPs, $z = 2.01$, $p < .05$. Thus, the expert groups agree more with each other than either group does with novices. Overall, the ratings sucessfully captured several important features of the subjects. The ratings reflect the differences among the three groups with intracorrelations being higher than intercorrelations; they reflect a difference between the student group and the expert groups first in lower consistency within the student group and, second, in less agreement between students and experts than between different groups of experts.

IV. MEASUREMENT OF CONCEPTUAL STRUCTURE

The goal of scaling procedures is to uncover latent structure in data. The structure is masked by the noise found in any set of errorful data. The scaling procedures assume that the latent structure obeys the assumptions of metric data, regardless of whether the empirical judgments meet these assumptions. For example, the latent structure is assumed to obey the triangle inequality assumption even though the empirical data contain violations of this assumption. In fact, the scaling procedures either manipulate the data to meet these assumptions, or they extract the parts of the data that meet these assumptions. Underlying these approaches is the belief that violations of the assumptions are due to noise and not to any meaningful psychological property (but see Tversky, 1977).

The differences among scaling procedures usually lie in the products of the procedures. There are procedures for generating hierarchical clusters, additive clusters, weighted free trees, multidimensional spaces, and general weighted networks (GWNs). Here we focus on the latter two procedures. MDS is a procedure that produces spatial configurations and has undergone conceptual, mathematical, and empirical scrutiny in a number of studies. Only recently GWNs have been derived from empirical dissimilarity data; however, they complement the spatial procedure in a number of ways and have distinct advantages over other nonspatial scaling procedures.

Methods

Multidimensional scaling. MDS is a powerful technique for extracting the latent structure within the errorful empirical similarity judgments. This is accomplished by arranging the concepts in N-dimensional space where the Euclidean distances between points reflect the psychological proximity of the concepts.

The first step in obtaining such representations involves submitting the empirical ratings, along with the desired dimensionality, to an MDS algorithm. The optimal dimensionality can be arrived at through a variety of techniques. One approach has been to generate several representations in different dimensionalities. The optimal dimensionality can then be selected by looking at stress (badness of fit) and $r2$ values. Once the data and desired dimensionality have been entered, the MDS algorithm returns a set of coordinates corresponding to the location of each concept in the space. The final step involves interpreting the resultant space along with the accompanying dimensions.

The MDS supplies several important pieces of information. First, it summarizes the data into a spatial configuration, which is complex at times, but is considerably more informative than are the empirical similarity judgments. Second, MDS captures the *global*

relations among the concepts. That is, MDS considers the relationship of each concept to all other concepts and places the concepts along the dimensions of the space in a way that reflects these relations. Although such a procedure can distort local relationships (that is, the distance between any particular pair), the procedure is unsurpassed at revealing global structure. In particular, successful identification of the dimensions of the space supplies information about conceptual structure that cannot be gleaned from the original ratings nor from other scaling techniques. Finally, MDS supplies a metric (distance between concepts in multidimensional space) that has some useful applications.

Although a concept can be located in multidimensional space by a series of coordinates (one for each dimension), the coordinates are not as useful as are distances in comparing representations. The distance between a pair in MDS is based on the Euclidean distance between the two points located by the coordinates corresponding to each concept. Distances in MDS preserve the structure of the representation, are independent of dimensional rotation, and meet the three standard metric assumptions: identity, nonnegativity, and triangle inequality.

To illustrate MDS, consider the two-dimensional spatial configuration in Figure 1a. The solution was based on the pairwise similarity judgments of undergraduate psychology majors for 16 naturally occurring objects. MDS has summarized the 16 x 16 matrix of distances into a representation that allows consideration of the global relations among concepts. In particular, the horizontal dimension reflects a nonliving-living dimension; the vertical dimension is more difficult to label but seems to have captured a difference between plants and animals, although this applies to only the living members of the space. In order to fix the concepts in space, MDS has introduced some local distortions. Maple is closer in space to rose than it is to tree. These local distortions are due to the adjusting of the data that occurs in order to produce distances that obey the metric assumptions. It will will seen that GWN complements MDS nicely by supplying information about local relations at the expense of global ones.

<u>General weighted networks</u>. A GWN is a configuration in which concepts are depicted by nodes, and relationships are depicted by links connecting the nodes. The links are assigned a value or weight that reflects the strength of the relationship between the nodes. The value reflects the distance from one node to another along that link; the shorter the link, the closer the nodes. The network is general in that constraints are not placed on the possible relations that can be represented. For example, the hierarchical constraint found in cluster analysis is not placed on GWNs. Without this constraint, the representation is free to contain local relations other than hierarchical ones, although hierarchical relations may still be present (Christofides, 1975; Fillenbaum & Rapaport, 1971).

Figure 1. (a) Two-dimensional MDS solution for 16 natural concepts.

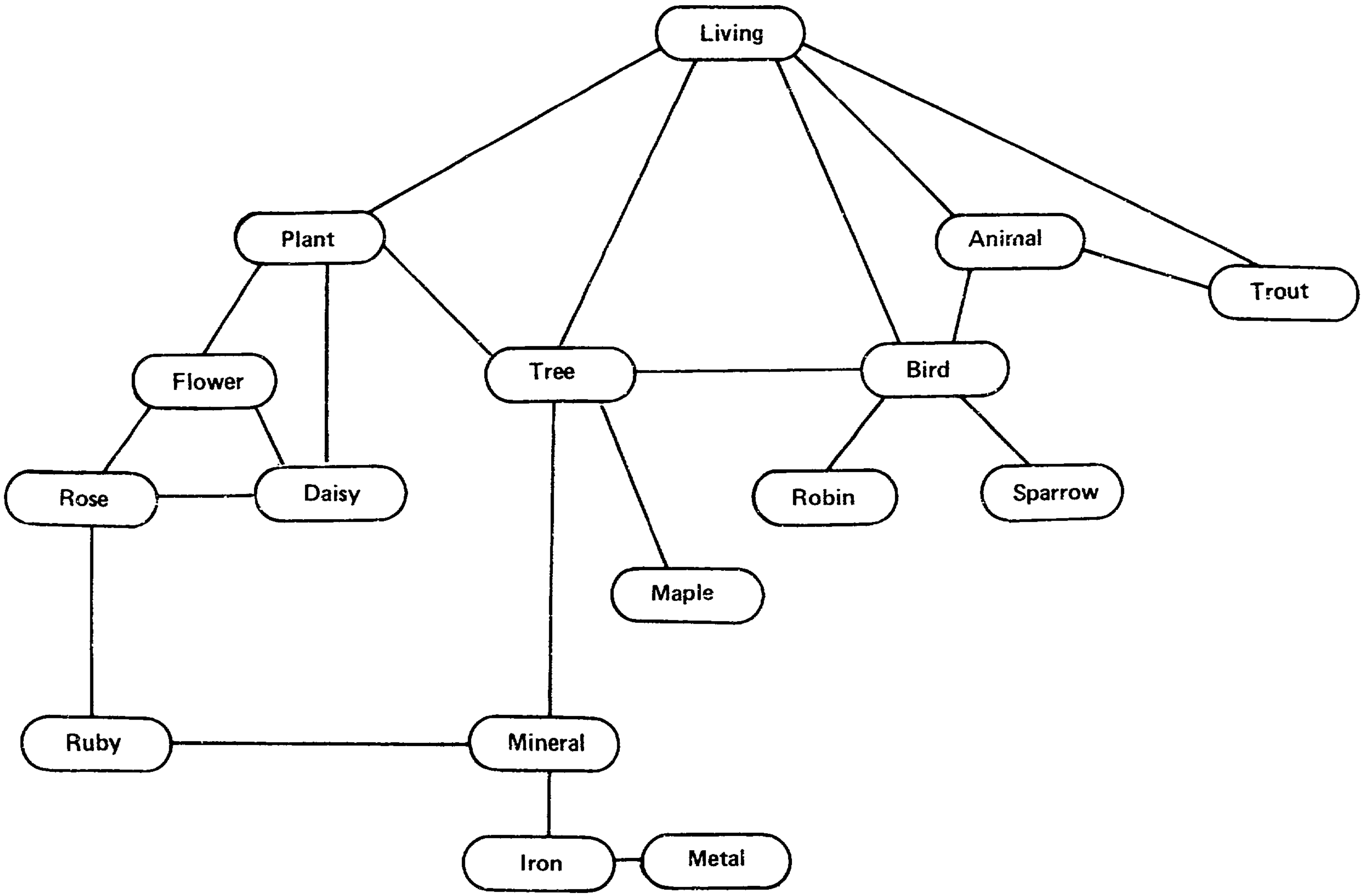

Figure 1. (b) GWN solution for the same concepts. Link weights have been omitted.

As has been noted, networks have formed the basis of research in a number of areas of cognitive science. Several psychological and artificial intelligence models of conceptual structure are based on such networks. The area of mathematics called graph theory is centrally concerned with properties of general networks. Although important theoretical and formal work has been conducted on these structures, no methods have been available until recently to produce networks from empirically obtained measures of psychological distance. GWN, an algorithm that produces general weighted networks, was applied to the rating data.

The central problem in constructing a network is to determine which links to place in the network. For N concepts, the possible number of links lies between N-1 links for a minimally connected network and N taken two-at-a-time links for all possible connections. The resulting network should obey the standard metric assumptions. GWN is a solution to the problem. It extracts information from the data set that obeys standard metric assumptions, unlike MDS which transforms the data to meet the standard metric assumptions.

The GWN algorithm is best illustrated by the natural concepts used to illustrate the results of MDS. GWN adds a link to the existing network for any two concepts, say, daisy and maple, in the following way. The empirical distance between the two concepts is compared with the length of the shortest chain (sequence of links) already existing in the network connecting daisy and maple. Such a chain might be daisy-plant-tree-maple, or more generally, daisy-X1-X2- . . .- Xi- maple, where the Xs represent intervening nodes of the chain. If the empirical dissimilarity is larger than the length of the shortest chain currently in the network, then GWN does not add a link because the new link would be redundant with the shortest chain; that is, a chain exists which can account for the empirical data. If the empirical distance is shorter than the evaluated distance of the shortest current chain, then GWN adds a link connecting daisy and maple (in the present example) because the psychological distance is smaller than would be allowed by the existing network; that is, GWN assumes that a chain cannot account for an empirical judgment that is less than what currently exists in the network. By iterating this procedure starting with the smallest empirical distance and proceeding with all distances in order of their magnitude, GWN adds links to the network and can create networks of varying complexity. The final result is that GWN produces a network in which a link is present if and only if the link is a necessary link in some minimum chain connecting two concepts. Thus, the network is the union of all possible minimum chains.

In the actual solution shown in Figure 1b, maple and daisy are not linked. A chain already existed in the network which could have accounted for the empirical similarity judgment of maple-daisy and so a direct link was not added. Another way to view this is that the

link between maple and daisy was not involved in the shortest path between any two concepts in the network. Notice that maple and daisy were close according to MDS but were not linked by GWN. Because GWN extracts the latent structure rather than transforming the data, it is better able to reflect psychological proximity on a pairwise basis. On the other hand, GWN does not produce global information of the kind supplied by MDS.

The use of general networks and the GWN algorithm holds substantial promise in attempts to specify the local relations and structure present in a conceptual organization. In addition, it allows a detailed, concept-by-concept comparison across groups that differ in expertise.

Results and Discussion

MDS. In this section, the properties of MDS are discussed. The work will focus on instructor pilots. Three dimensions supplied the optimal dimensionality for both the split-plane concepts and the low-angle strafe concepts for all groups of pilots. Inspection of r2 and stress as a function of dimensionality suggested optimal dimensionality of 3 or 4. The method of Isaac and Poor (1974) suggested that three-dimensional solutions were more appropriate than were four. The strafe scenario, because it is inherently less complex, had been expected to yield a smaller dimensionality. Perhaps when subjects are required to consider a single maneuver, nuances in the maneuver receive more attention than would be the case from a more global point of view. In addition, it was somewhat surprising that the novices (UPs) and experts (IPs and GPs) each resulted in a solution of equivalent dimensionality. In the later discussion of GWN, it will be suggested that this may be a limitation of the MDS scaling procedure used here, rather than a suggestion of equally complex solutions for novices and experts.

The results of three-dimensional scaling solutions of split-plane maneuvers for IPs are shown in Figure 2. Figure 2a presents the position of each concept along the first two dimensions, Figure 2b presents the position of each concept along the first and third dimensions, and Figure 2c presents the position of each concept along the second and third dimensions.

In order to identify the dimensions, assistance was obtained from personnel at Holloman AFB and Williams AFB. Each of the dimensions has been identified for the split-plane solution, and one of the dimensions has been identified for the strafe solution. The split-plane concepts have one dimension associated with a temporal factor, one dimension which distinguishes particular maneuvers, and one dimension associated with factors distinguishing concepts that are related to distance from concepts related to attitude. The temporal dimension identifies the general time dimension within a scenario leading to split-plane maneuvers. In Figures 2a and 2b,

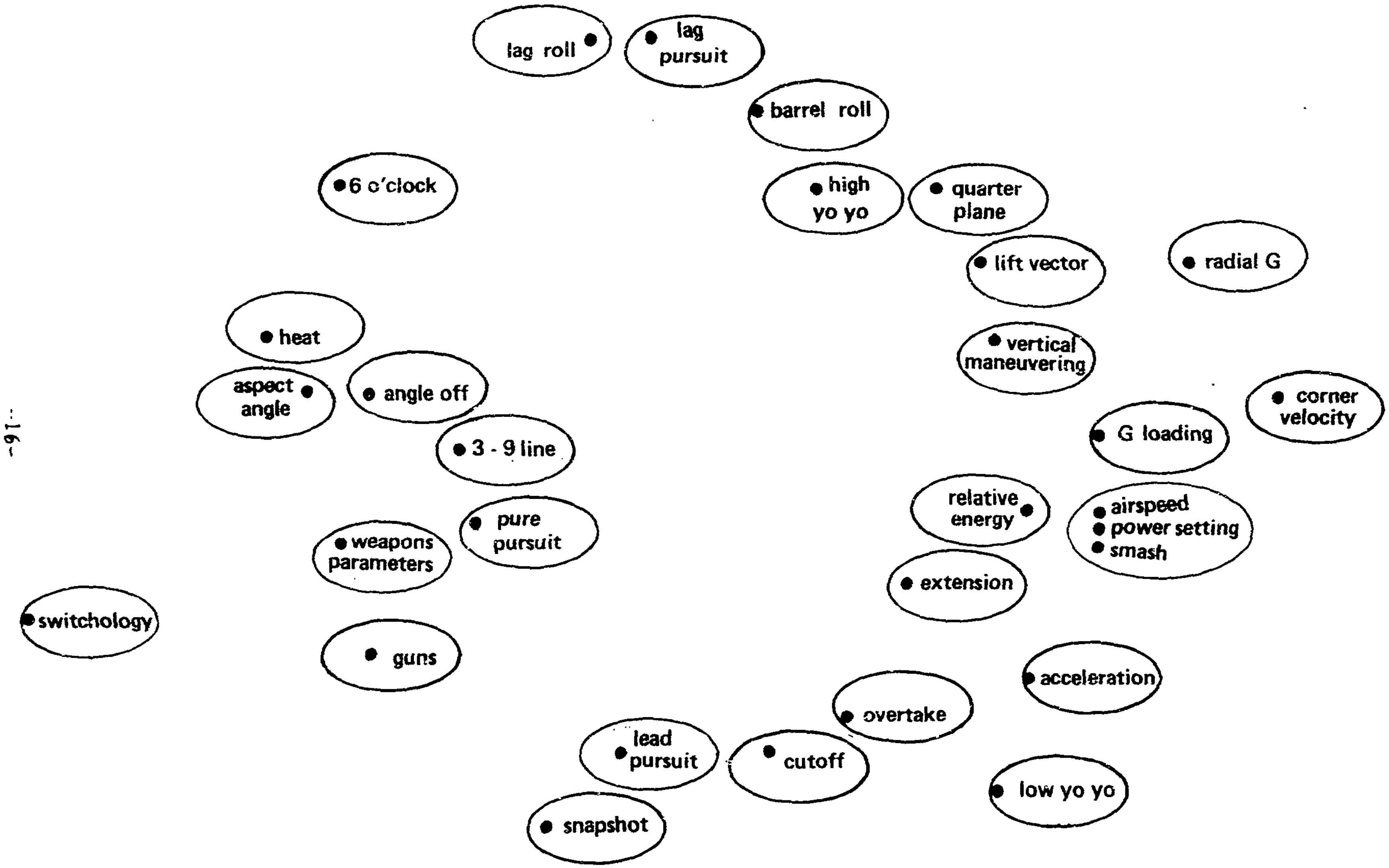

Figure 2. (a) Dimensions 1 (horizontal) and 2 (vertical) of a three-dimensional MDS solution for instructor pilots in the split-plane scenario.

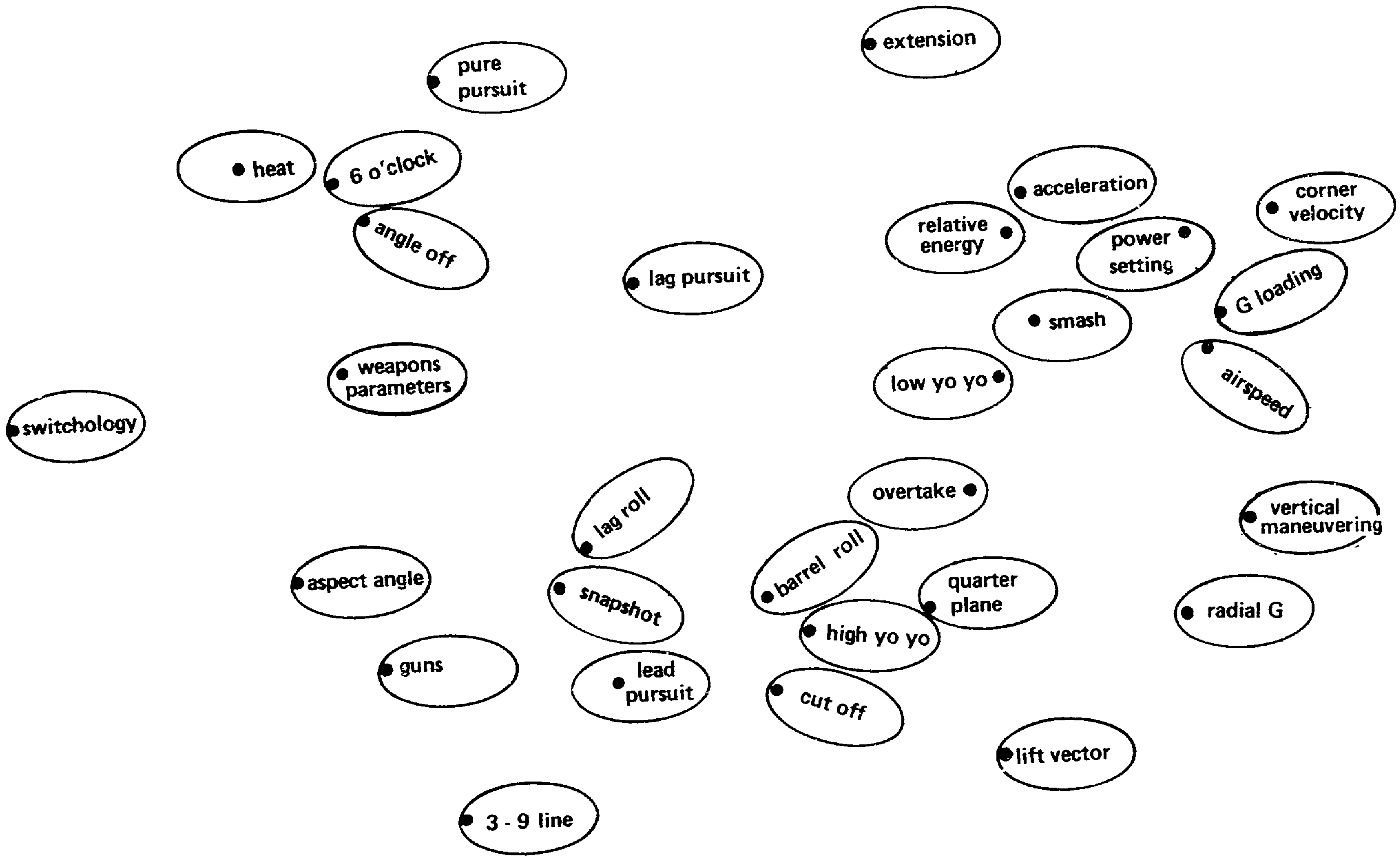

Figure 2. (b) Dimensions 1 (horizontal) and 3 (vertical).

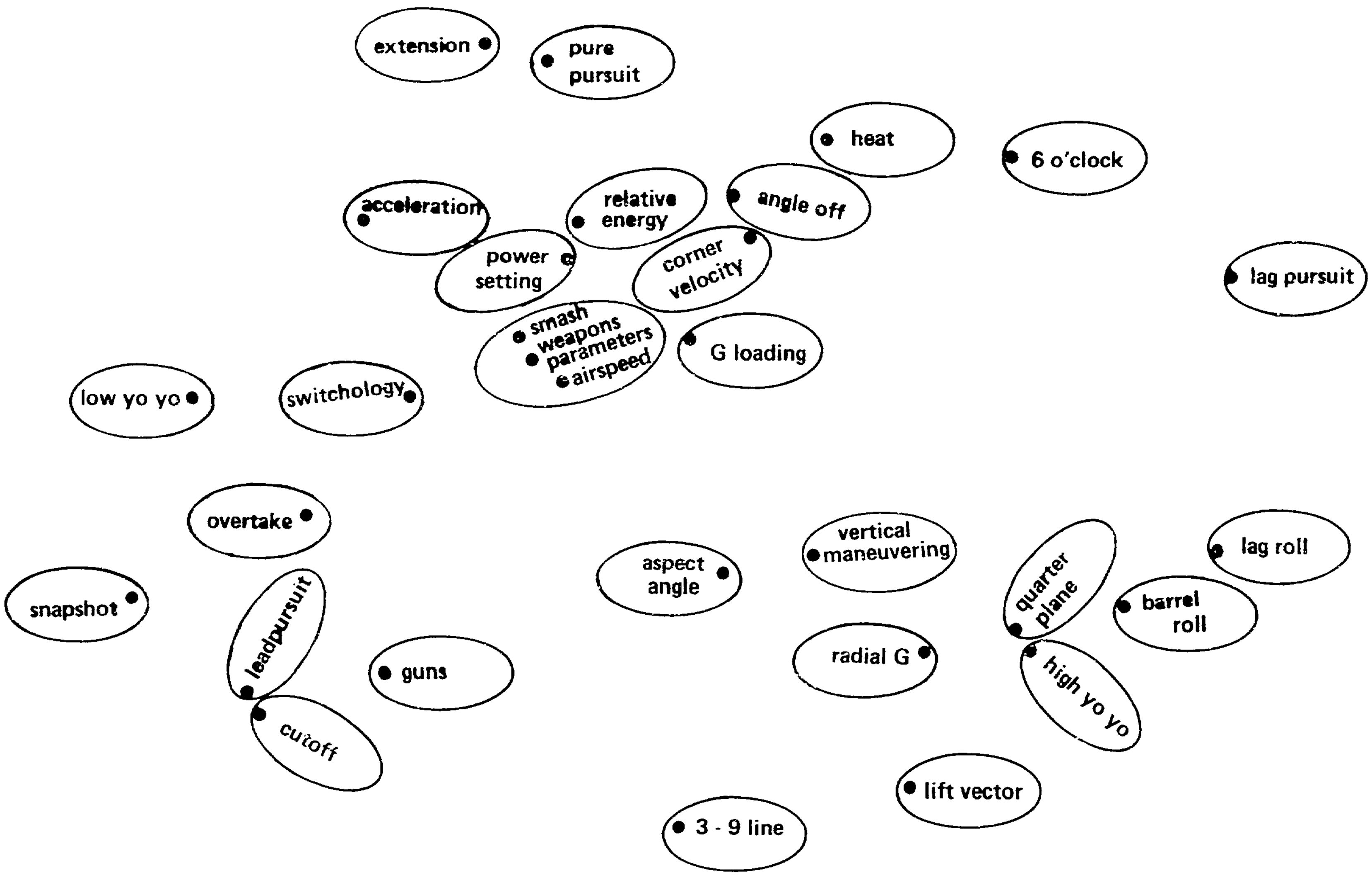

Figure 2. (c) Dimensions 2 (horizontal) and 3 (vertical).

this dimension is the horizontal dimension ordered from left to right. The concepts on the extreme left (SWITCHOLOGY, HEAT, and ANGLE-OFF) refer to events or considerations that occur early in the temporal sequence. To the right, the concepts refer to events and considerations occurring later in the sequence. Concepts occurring later in the sequence are actually consequences of actions performed early in the sequence. The second dimension is a contrast between lead pursuit and lag pursuit with LAG PURSUIT and the associated maneuvers near the top and LEAD PURSUIT and LOW YO YO near the bottom in Figure 2a. This dimension is again represented along the horizontal dimension in Figure 2c. The third dimension is the vertical dimension in Figures 2b and 2c. This dimension has been tentatively identified as separating concepts that refer to actions and considerations related to the range or distance between aircraft from concepts related to the relative positions of the aircraft. This dimension separates concepts that concern distance from concepts that concern attitude.

The low-angle strafe maneuver also yielded a temporal order dimension in the solution. Again, this dimension occurred as the first dimension in the solution, and it reflects the order in which the concepts would occur to pilots in executing the low-angle strafe. Interestingly, these temporal dimensions appear to reflect the psychological ordering of the concepts rather than the order in which the events occur in physical time. Apparently, pilots must consider several factors early in time, before they actually occur, in order to be able to concentrate on critical factors such as aiming and firing. The temporal order dimension is a powerful one in the organization of these concepts for pilots.

The dimensional organization of the concepts is interesting, and it lends some tentative support to the validity of the analytic procedures underlying the MDS solutions. More fine-grained analyses of the structures are required to lead to conclusions that may be usefully applied. The metric-based analyses considered in the section of this report on validation represent a step in that direction.

GWN. An analysis was performed on the data from UPs, IPs, and GPs for split-plane maneuvers and the low-angle strafe maneuver using GWN. The resulting networks for IPs and UPs appear in Figures 3 and 4. The nodes in the networks are located on the page according to the two-dimensional MDS solutions for the IPs. One of the problems of representing the networks is arranging the nodes. Using the MDS solution solves that problem and has the advantage of depicting both dimensional information and network information in the same representation. The MDS solution for the IPs is used for both networks to facilitate comparisons between IP and UP networks.

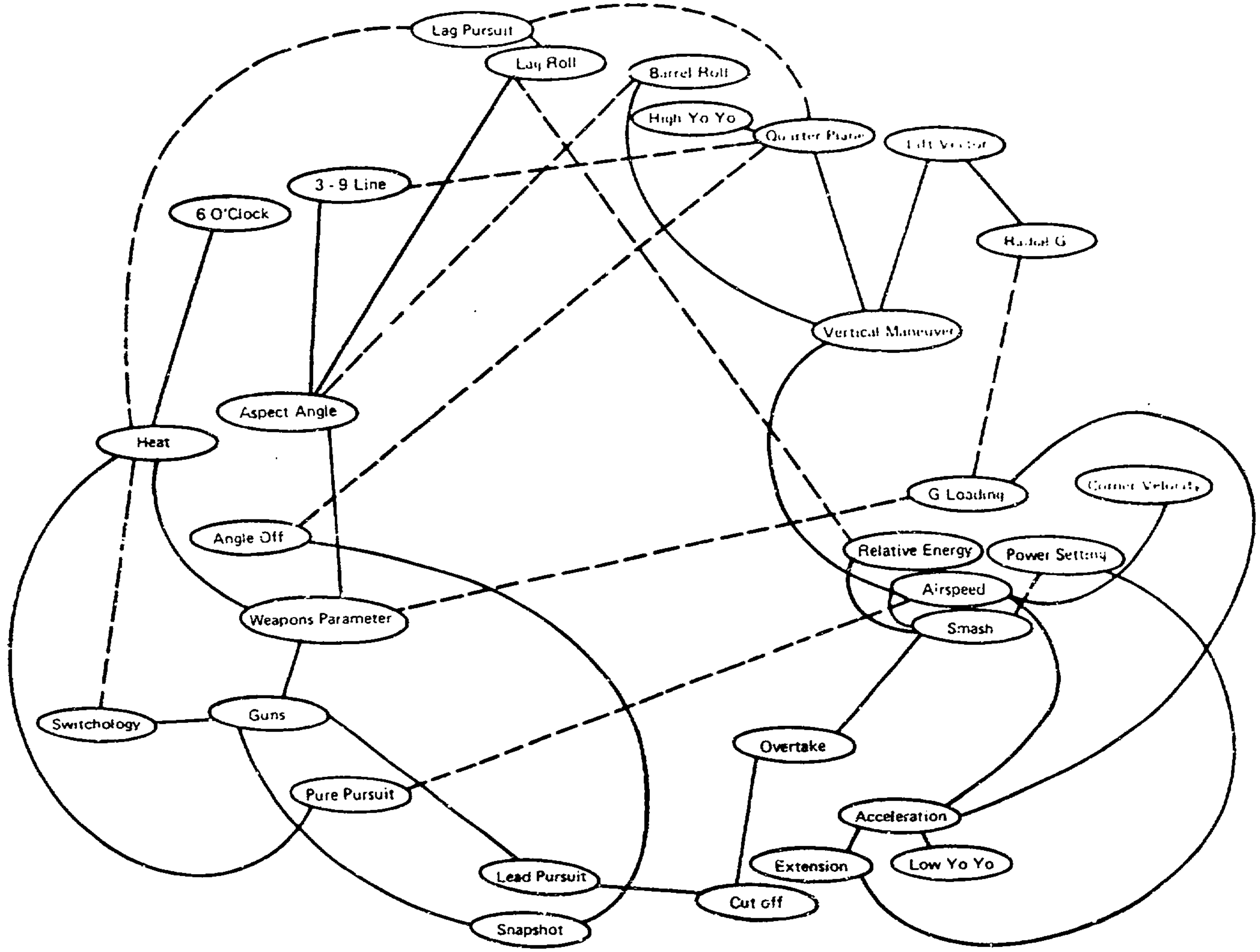

Figure 3. GWN solution superimposed on a two-dimensional MDS solution for instructor pilots in the split-plane scenario.

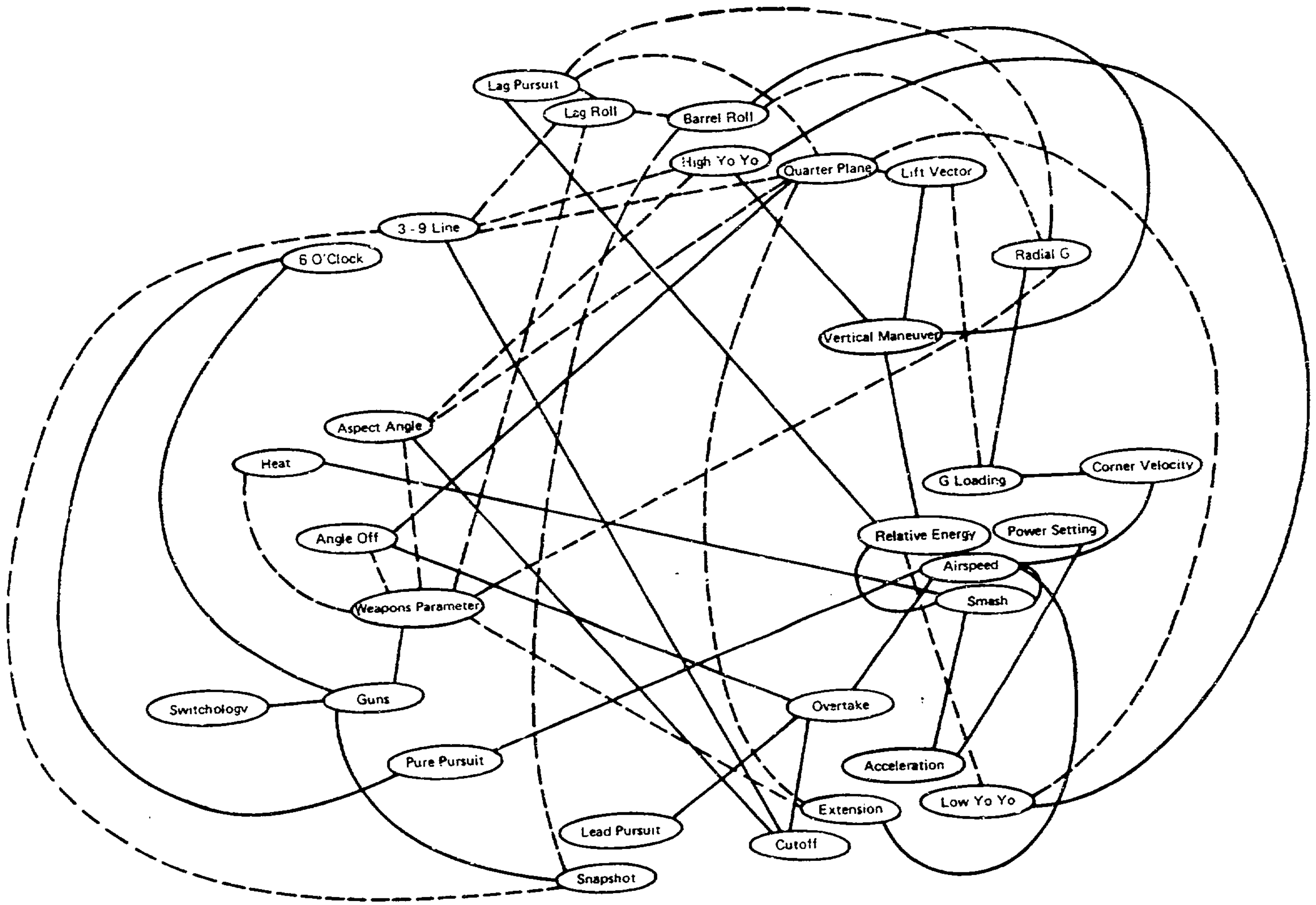

Figure 4. GWN solution superimposed on a two-dimensional MDS solution for undergraduate pilots in the split-plane scenario. The MDS solution is that of IPs.

Consistent with the MDS analyses, GWN supplied networks of comparable complexity for the split-plane and the strafe scenarios, disconfirming the expectation that the strafe maneuver would be viewed as less complex. However, GWN does suggest that conceptual structures of experts and novices do differ in complexity. The most striking difference between the IP and UP networks is that the network derived from the student data is considerably more complex than the IP network. The UP network has 51 links compared to 40 for the instructor network. This pattern is even more extreme for the strafe concepts, with IPs producing a structure of 37 links compared to 65 links for the UPs. This result can be contrasted with MDS which yielded three-dimensional solutions for all pilot groups. Apparently a characteristic of expertise is not a more complex structure, but rather, experts tend to identify the important, critical information and associations, yielding a simpler network.

The weights of the links (not shown) varied to an extent, with an average link length of 10.4 and a standard deviation of 7. The shortest link is between GUNS and SNAPSHOT (.1 unit), and the longest links are LAG ROLL-ASPECT ANGLE and HEAT-PURE PURSUIT (2.6 units). Interestingly, no relationship that received a dissimilarity rating greater than 2.6 manifested itself as a link in the network. The links in the network, including the long connection between HEAT and PURE PURSUIT, all represent connections between related concepts.

GWN also reveals other interesting structural facets of the conceptual structure of the IPs. For example, there are several concepts that link to multiple concepts (GUNS, HEAT, ASPECT ANGLE, VERTICAL MANEUVERING, ACCELERATION, SMASH, AIRSPEED) suggesting that the representation is not strictly hierarchical. In addition,the network of split-plane concepts for the IPs highlights a number of local relationships that are not apparent in the MDS scaling techniques that were used here.

Summary. Both MDS and GWN reduce a large amount of data in the form of a dissimilarity matrix to a much smaller set of data, but they tend to highlight different aspects of the underlying structure. GWN provides an understanding of the local relationships among concepts, whereas MDS provides a more global understanding of the dimensionalized concept space. In MDS's effort to find the best Euclidean fit to the data, it has sacrificed some of the pairwise distance information. That is, it has placed concepts near each other in space that were not viewed as related, and it has separated related concepts. Because the network extracts information from the rating data that follows the metric assumptions rather than altering the rating data to meet the metric assumptions, the links present in the network highlight pairwise distance information. On the other hand, GWN does not supply the type of global information that allowed the identification of the underlying dimensions of the conceptual space, as did MDS. Inspection of the results for IPs can reveal some of the similarities and differences of MDS and GWN.

V. VALIDATION

The previous sections have described techniques for representing conceptual structures. MDS and GWN both produce relational and organizational information about concepts within a particular domain of knowledge. One means of validating and comparing MDS and GWN is to use these conceptual structures to discriminate among groups and to predict group membership. It is reasonable to assume that differences in experience among the pilots should be mirrored in differences among their conceptual structures. It is also reasonable to assume that members of a group of pilots share certain characteristics in their conceptual structures. Given an IP's conceptual structure for the split-plane maneuvers, can this individual be correctly identified as an IP? Accurate classification based on conceptual structures would support the validity of the structure. Classification also provides one means for comparing different ways of defining conceptual structures. Furthermore, classification procedures provide a means for assessing the nature of group and individual differences. Thus, the interest in this phase of the project is in evaluating the validity of conceptual structures and in assessing similarities and differences of structures both across and within groups of pilots.

Discrimination. In addition to arranging a set of concepts in multidimensional space, scaling techniques are available for placing individuals or groups of individuals in multidimensional space. Carroll and Chang's (1970) individual differences MDS can be used to locate individuals along the same dimensions in which concepts are placed. Thus, the output of this procedure illustrates which dimensions are most highly weighted for specific groups or individuals. For instance, a point located close to the zero coordinate for a particular dimension indicates this dimension was not critical for that particular group. The Alscal version of Carroll and Chang's INDSCAL program is used here.

The location of groups of pilots along the three dimensions used previously in this work is not only useful in further distinguishing groups from each other, but is also a valuable technique for validating the dimensions. If the resulting three dimensions are meaningful, it is expected that these dimensions would be more critical to expert pilots than to the novices who lack the understanding and organization of the concepts found in more experienced pilots.

Method & Results. In order to plot the three groups of pilots in multidimensional space, the distance matrices for each of the IPs and each of the GPs for each scenario were submitted to an individual differences scaling procedure. This yielded an expert space. The dimensions found earlier with classical MDS were mirrored in the INDSCAL solution. The distance matrix from each UP, one at a time, was added to the distance matrices of the experts, and then the space

Table 3. Weighting of Dimensions for Each Group of Pilots for Split-Plane and Strafe Scenarios

a. Split Plane Manuevers

	Dimension 1	Dimension 2	Dimension 3	
	Temporal	Distance-Attitude	Lead-Lag	Mean
IPs	.3475	.3142	.3270	.3296
GPs	.3023	.2856	.2717	.2865
UPs	.1923	.1779	.1761	.1821
Mean	.2807	.2592	.2583	

b. Strafe Maneuver

	Dimension 1	Dimension 2	Dimension 3	
	Temporal	Unknown	Unknown	Mean
IPs	.3819	.3327	.3020	.3389
UPs	.1844	.1833	.1687	.1788
Mean	.2832	.2580	.2354	

was recomputed. The dimension weights for each UP was recorded and compared to the dimension weights for the experts as derived from the expert space. Results shown in Table 3 indicated that UPs weight each dimension less than do experts. Planned comparisons setting experiment-wise alpha to .05 (test-wise alpha = .006) confirmed these conclusions in each case. For the split-plane scenario, experts relied more heavily on the temporal dimension, $t(32) = 7.84$, the attitude-distance dimension, $t(32) = 7.73$, and the lead-lag dimension, $t(32) = 6.98$, than did the novices. Similarly, in the strafe scenario, the temporal dimension was weighted more by experts, $t(21) = 10.33$, as were the two unidentified dimensions, $t(21) = 6.32$ & $t(21) = 8.52$, for dimensions 2 and 3, respectively.

As mentioned previously, heavier weighting of the dimensions by experts would be expected if the dimensions had a psychological validity; UPs lack the understanding of the dimensions and organizational strucure of the more experienced pilots. It appears that none of the dimensions are as salient for the UPs as for the more experienced groups.

Secondly, results of this scaling indicated that the two groups of experts are also discriminable: IPs tended to weight each dimension in the split-plane solution more than did GPs. This tendency was, however, only statistically reliable for the lead-lag dimension, $t(14)=2.25$, $p < .05$. Some differences between IPs and GPs would be expected given that the groups have dissimilar backgrounds. The initial qualified interpretation of these dimensions receives some support from the INDSCAL findings. The overall tendency for IPs to use the dimensions more than do GPs may reflect their classroom experience.

Classification

In the previous section, it was shown that the groups were discriminable based on their conceptual structures. Here, methods are developed for classifying an individual as a member of a particular group based on the individual's conceptual structure and then pattern classification techniques are used to analyze conceptual structures.

Classification procedures are generally concerned with the problem of assigning an object to one of two or more groups. The groups may vary along several attributes or variables. The groups are defined such that each object belongs to only one group. Objects are represented by a list of numerically described attributes. The general notion of classification involves comparing each object's position to each group's prototype in order to locate the closest group.

There are many different classification techniques. One common method is based on discriminant analysis. Discriminant analysis requires that a number of assumptions be met including that, (a) objects be measured at the interval or ratio level of measurement because discriminant functions are computed with means, variances, and correlations, (b) the number of objects exceed the number of attributes defining an object, (c) no attribute is a linear combination of other attributes, (d) population covariance matrices for each group are equal, and (e) group populations have multivariate normal distributions.

Because of problems meeting some of these assumptions, the present analyses used a classification procedure based simply on distances in feature space. Nilsson (1965) provides a general discussion of this technique. The major difference between this technique and discriminant analysis is that variances and correlations do not enter into calculating discriminant functions. With this technique, objects to be categorized are represented by a list of feature values in the form of a pattern vector. The ith element of the vector represents the value of the ith feature. Because feature values are in the form of real numbers, pattern vectors can be considered as points in a multidimensional space where each dimension represents an attribute of the object. The goal is to develop discriminant functions that will partition the pattern space into regions containing only those points or patterns belonging to a particular class of patterns.

Linear discriminant functions (the only type used here) assume that a weighted linear combination of the feature values can classify patterns. A linear discriminant function has the form:

$$g(X) = W1X1 + W2X2 + \ldots + WdXd + Wd + 1,$$

where W=W1,W2,...,Wd is a vector of weights. Classes that can be properly separated with linear discriminant functions are known as linearly separable. The first analysis consisted of determining a discriminant function to classify all but one person from each of two groups. An attempt was then made to classify the remaining two individuals. This procedure was repeated for all possible combinations of N-1 people from one group and N-1 people from a second group.

The method for generating linear discriminant functions here began with minimum distance classification. In this case, a prototype point representing the central tendency of a class of patterns is constructed for each group of N-1 individuals. The prototype is simply the average of the feature values of all patterns belonging to a group. The initial linear discriminant function, or the decision surface separating the patterns, was the perpendicular bisector of a line connecting the two prototype points.

If this initial function successfully classifies all the N(1)-1 and N(2)-1 "known" individuals, it then stops and makes a judgment classifying the two "unknown" individuals. If, however, the starting function fails to classify correctly the training set of known individuals, a training procedure alters the function by successive adjustments to the weight vector W by adding a fraction of the pattern vector that was incorrectly classified to the weight vector. This produces a new weight vector $W' = W + cX$, where c is a positive number that controls the extent of the adjustment. The procedure is terminated as soon as the weight vector correctly classifies all patterns in the training set. The remaining two individuals are then classified.

These pattern classification techniques can provide one measure against which to test the validity of representations of conceptual structure. For example, assuming that there are differences in the way novices and experts organize knowledge about the area in question, empirical methods of representing memory organization should reflect this difference in expertise. Pattern classification techniques can then be used on information derived from empirically generated representations of conceptual structure, in an effort to classify individuals as novices or experts.

Method. Pattern classification analysis was performed on data obtained from GPs, IPs, and UPs for split-plane maneuvers, and on IPs and UPs for low-angle strafe maneuvers. Three types of patterns were generated for each individual, based on MDS, networks, and raw ratings. Network patterns were formed for each individual by taking the presence or absence of links in the network for each pair of concepts. The network pattern for each subject consisted of a vector of ones and zeroes representing the presence or absence of a link, respectively, between each pair of concepts. An MDS pattern consisted of a vector of the distances between the members of each pair of concepts in multidimensional space. Patterns based upon the original ratings were formed by considering the dissimilarity rating for each pair of concepts as a feature of a pattern. All three methods resulted in patterns with 435 features corresponding to all the possible pairs of 30 concepts.

If GWN and MDS successfully capture the latent structure within the ratings, classification based on these techniques should be superior to classification based on the ratings. Further, classification of the experts (i.e., IPs and GPs) should be more difficult than any classification involving UPs, if these techniques have produced psychologically valid measures of conceptual structure.

Results & Discussion. The results of the pattern classification analysis are shown in Figure 5. These percentages are based on different numbers of classifications, depending on the number of members of each group. In general, there are 2 x (N1 x N2) individual classifications for each comparison. Thus, the number of

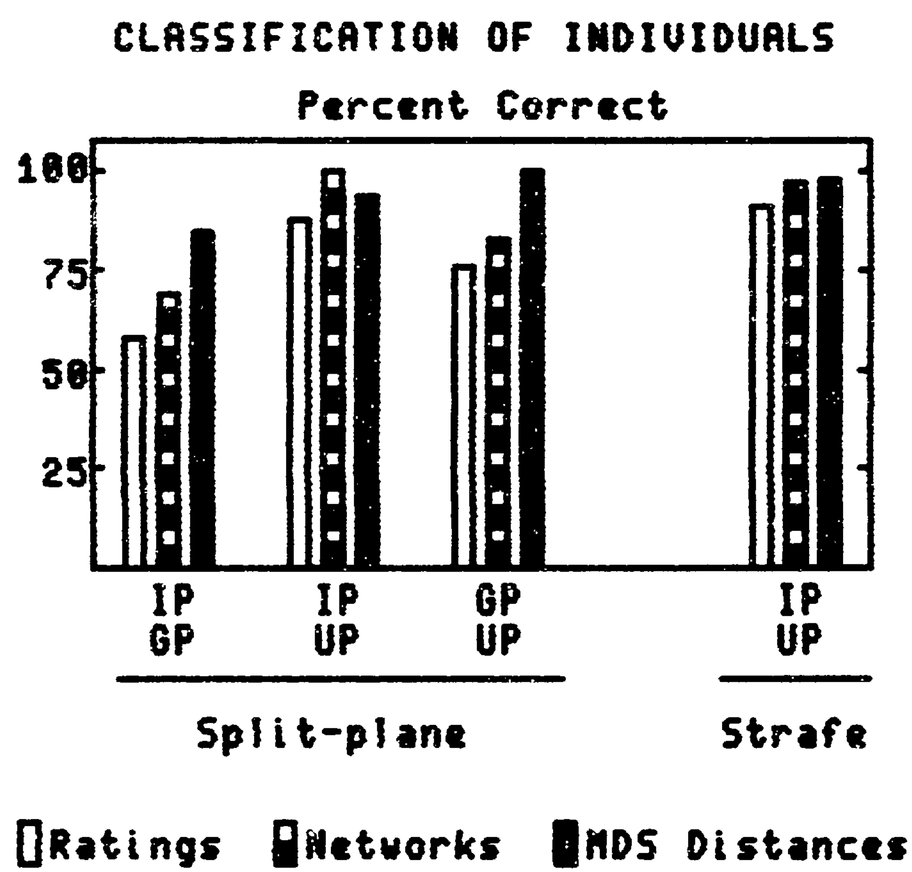

Figure 5. Results of pattern classification procedures applied to each pair of pilot groups for both scenarios.

possible correct classifications ranged from 306 for UP-GP down to 126 for IP-GP. Overall, classification of individuals into groups was better than chance for patterns derived using all three methods, especially in discriminating novices (UPs) from experts (IPs and GPs).

Pattern classification using the original ratings was least successful, classifying 85% of the novices and experts and 59% of the experts. Pattern classification using distances derived from MDS was most successful, with a mean classification performance of 97% on novices and experts and 85% accuracy in discriminating GPs from IPs. For each case, classification based on MDS distances was superior to classification based on the original ratings. MDS was very successful at finding the underlying structure in the rating data. Classification accuracy using patterns derived from networks was also better than the original ratings in each case. Network patterns classified 93% of the experts and novices correctly on the average and correctly classified 69% of the IPs and GPs. Networks also appear to supply information not apparent in the original ratings that is useful in classifying pilots.

Several conclusions follow from these results. First, pattern classification techniques can discriminate novices from experts with a high degree of success based upon both structural descriptions of their conceptual structure, and raw similarity ratings. Second, both MDS and GWN showed superior classification compared to the empirical ratings, with MDS outperforming GWN by 4% on expert-novice classifications and 16% on expert-expert classifications. Finally, all procedures found it more difficult to make expert-expert classifications based on conceptual structure than expert-novice classification.

It is interesting that the network patterns classified so well, especially given that the vectors for networks, unlike MDS or the ratings, were comprised only of a list of zeroes and ones. For example, 100% of the IPs and UPs were classified correctly simply by knowing which concepts were linked in the networks. On the surface, this may seem like the loss of a considerable amount of information; however, this loss of information is apparently compensated for by the network's ability to uncover the latent structure. It was suspected that the performance of the network may be improved, relative to MDS, if we equated MDS and GWN on the amount of relative to MDS, if MDS and GWN were equated on the amount of information contained in their respective patterns.

<u>Method</u> & <u>Results</u>. One way in which MDS vectors and GWN vectors can be equated for purposes of classification is to make the MDS vectors similar to the GWN vectors by transforming the MDS vectors to patterns of zeroes and ones. MDS distances were converted to patterns containing only zeroes and ones. In order to produce patterns that were as comparable to the network patterns as possible,

an effort was made to produce patterns with roughly the same number of ones and zeroes as are contained in the network patterns. A cutoff was selected based on the number of links in each corresponding network. All pairs of concepts with MDS distances at or below this cutoff were assigned a 1, and the remaining concepts were assigned a 0.

Classification using 0/1 patterns converted from MDS distances was successful in 95% of the expert-novice cases and 87% for the expert-expert classifications. This compares with the MDS classification using all of the metric power of MDS (97% & 85%). Thus, when MDS distances are reduced to a pattern of zeroes and ones, classification remains superior to classification based on the networks.

These results suggest that the structures imposed by MDS and networks on raw data capture some valid structural information about the differences in the way distinct groups of pilots organize conceptual information. In comparing MDS to network representations, it would appear that MDS captures somewhat more of the structural information that is useful in discriminating between groups of individuals. The MDS superiority was especially noticeable in discriminating between the groups of experts.

Summary

In this section, an attempt was made to validate the conceptual structures obtained in the previous section. Validation took two forms. The primary form of validation consisted of attempts to recover differences among the pilots, especially between novices and experts, based on their conceptual structures. This proved quite successful in that the groups were discriminable and the individuals were classifiable. The secondary vehicle of validation showed that conceptual structures based on MDS or GWN contained more useful information than did the original ratings.

VI. POSSIBLE APPLICATIONS

Prediction and selection

Prediction of pilot performance and selection of pilot trainees for job placement are two important potential applications of this work. Knowledge concerning the differences between each individual pilot trainee and a group of expert pilots enables one to select the single trainee who most resembles the experts. Information of this sort may also allow prediction of performance based on the individual's previous standing with other members of the group. There are several techniques that serve the purpose of identifying individual differences. The methods described in the following paragraphs have each been used previously in this work to achieve other goals. For instance, pattern classification has been used as a means of validating various types of cognitive representation (networks, MDS). However, this technique can also be used to examine the similarities or differences of an individual in relation to his group or to other groups. The assumption that individuals and groups differ in their cognitive organization of the selected fighter pilot terms underlies each of these techniques.

The use of these techniques is exemplified with the split-plane scenario. In each case, an attempt was made to rank the UPs in relation to the experts. Following a discussion of each technique, the different rankings of the UPs are compared using the different techniques.

Pattern Classification. Pattern classification techniques are useful in reflecting how similar an individual is to his group and how similar an individual is to some other group. Two measures are of particular interest, the distance from an individual to the other group's prototype (how like the other group is the individual) and the distance from an individual to the belonging group prototype point (how like the group is the individual). The prototype point is that point representing the central tendency of a class of patterns. The distance between an individual and a class prototype reveals how strongly that individual represents the average features of that class. Because MDS distances yielded the best overall classification, MDS is used here to consider individual differences. When network patterns of zeros and ones were used in pattern classification, the classification was perfect (as it was with MDS distances). The ranks of the UPs were identical with MDS and network patterns.

Table 4 (column 1a) gives the distances from an individual to the group prototype for experts (IPs and GPs) and (column 1b) the distance from each UP to the group prototype for novices (UPs). All individuals were classified on the correct side of the decision surface. Next to each UP score is a rank indicating how closely the UP is to the expert prototype, with a rank of 1 indicating similarity to the experts.

Table 4. Individual Difference Data for Split-Plane Concepts (Numbers in Parentheses are Ranks for the UPs. Low Ranks (1) Suggest High Similarity to Experts.)

	Pattern Classification		INDSCAL	Person Space
	Distance to Expert Prototype		Distance from Origin	Distance from Origin
IP1	110		.530	208
IP2	145		.437	190
IP3	106		.592	214
IP4	130		.545	207
IP5	122		.595	208
IP6	142		.458	170
IP7	103		.631	227
GP1	125		.458	224
GP2	133		.481	167
GP3	111		.575	209
GP4	132		.515	241
GP5	129		.448	238
GP6	140		.429	189
GP7	122		.525	196
GP8	173		.354	000
GP9	116		.565	166
		Distance to Novice Prototype		
UP1	169 (6)	132	.367 (8)	346 (6)
UP2	180 (14)	149	.268 (16)	389 (13)
UP3	166 (5)	144	.482 (1)	263 (3)
UP4	193 (17)	162	.332 (10)	473 (17)
UP5	189 (16)	149	.267 (17)	441 (15)
UP6	159 (2)	141	.353 (9)	258 (1)
UP7	166 (4)	148	.471 (2)	318 (5)
UP8	173 (10)	145	.323 (11)	285 (4)
UP9	162 (3)	122a	.383 (5)	327 (6)
UP10	174 (11)	147	.310 (12)	353 (10)
UP11	170 (7)	135	.461 (3)	347 (8)
UP12	188 (15)	159	.298 (13)	447 (16)
UP13	171 (8)	128	.372 (7)	349 (9)
UP14	172 (9)	131	.447 (4)	356 (11)
UP15	177 (12)	156	.284 (14)	379 (12)
UP16	155 (1)	145	.380 (6)	259 (2)
UP17	179 (13)	144	.276 (15)	401 (14)

a prototypical UP

This information can be used to select individuals who are most like members of their own group or most like members of another group. For instance, UP16 is the UP closest to the expert prototype (rank of 1). This suggests that the conceptual structure of UP16 is most similar to that of the experts. Thus, one might predict that UP16 would perform more like an expert on a particular flight-related task than would other UPs. On the other hand, UP4 is furthest from the expert prototype, suggesting a larger distinction between this individual and experts than between UP16 and experts. Thus, these measures provide a means of detecting within and between group differences and, consequently, are prediction and selection aids.

UP9 is the closest to the prototype of the undergraduates. In fact, if pattern classification is conducted between experts and UPs, using only UP9 as a single "known" UP, classification is successful for 91% of the cases. Having this information about members of a class could have pedagogical value. Instructors are often concerned about the level at which to pitch a lecture. One possibility is to make certain the prototypical student has understood the material.

Individual Differences Scaling. Another method of ordering individuals is the INDSCAL MDS procedure. Earlier, this technique was discussed to show that the pilot groups were separable based on their cognitive structures. It is also possible to determine how each individual pilot weights the dimensions and then to compare individuals in the resultant space. INDSCAL locates individuals along the same dimensions on which concepts are located. Thus information provided by this scaling procedure e[illegible]ails how much the particular individual relies on a particular dimens[illegible]n. Earlier in this report, it was shown that UPs as a group tended [illegible]o weight the dimensions less than did experts. Here, the extent to which each individual UP considers the dimensions can be determined. Individuals can be ranked by their distance from the origin (0,0,0). For example, a UP that does not weight any of the dimensions would be 0 units from the origin and quite unlike any expert (who relies on the dimensions). A UP that weights the dimensions heavily has in some sense a conceptual structure more like that of an expert.

Table 4 (column 2) gives, for each individual, the distance from the origin of the three-dimensional MDS solution reported earlier. UP3 is the furthest from the origin of all the UPs, suggesting that this individual relied heavily on the same dimensions as did the experts, in making similarity judgments. In contrast, UP5 is very near the origin, suggesting that this UP made very little use of the expert dimensions. As with the pattern classification discussed previously, the INDSCAL procedure allows comparisons of UPs with experts. It has the advantage of restricting comparison to the same conceptual space for experts and novices, but does not supply the information about the prototypical student that the pattern classification technique yields.

Person space. The final individual difference technique to be considered here is a hybrid of the previous two techniques. It is an attempt to represent the individuals in a multidimensional space, but within a space that has dimensions relevant to the subjects, not to the concepts. Thus, the plan is to position subjects in a space where the dimensions reflect differences among the subjects. If this is successful, one of the dimensions should be of expertise.

To represent individuals in multidimensional space, an inter-subject distance matrix was derived, similar to the interitem distance matrix derived for the concepts. Distances for this matrix were derived from distances from each individual three-dimensional MDS solution. In this case, the individual can be thought of as a point in "n-dimensional" space (n = 435 dimensions based on the 435 distances for all pairs of 30 terms). The distance between two individuals would take into account the difference in distance for each of the 435 pairs of points for the two individuals. These distances resulted in a matrix of distances with individuals as rows and columns. The entries in this matrix were simply the distance from one individual to another. These distance values were then scaled in multidimensional space using one and two dimensions.

Figure 6 shows the two-dimensional MDS solution. The experts and UPs are clearly linearly separable. One of the dimensions can be readily identified as a dimension of expertise, suggesting the technique has been successful at establishing a multidimensional space with dimensions that characterize the subjects, rather than the concepts. The second dimension is suggestive of a pedagogical-operational dimension, with the IPs and GPs tending to occupy different ends of that dimension.

A one-dimensional MDS solution should extract the single dimension that accounts for the most variance. The values along that dimension were transformed to set the first pilot at a coordinate of 0; these values appear in Table 4 (column 3). Comparison of these values indicates the relative distance from one individual to another. Thus, it seems that the one-dimensional solution ordered individuals along an expertise dimension. The creation of the "person space" clearly helps to define the separate groups of subjects and the locations of individuals in relation to the groups. The earlier finding that the more experienced pilots agree more with each other and have well defined conceptual structures is supported by Figure 6, in that expert groups tend to form tighter clusters than do UP .

Again, UPs that are close to the expert end of the continuum should have organizations of flight-related information similar to those of experts. For this technique, UP6 appears closest to the experts, with UP4 being most distant. In general, representation of individual pilots in multidimensional space provides a measure of distance that can be used in prediction and selection.

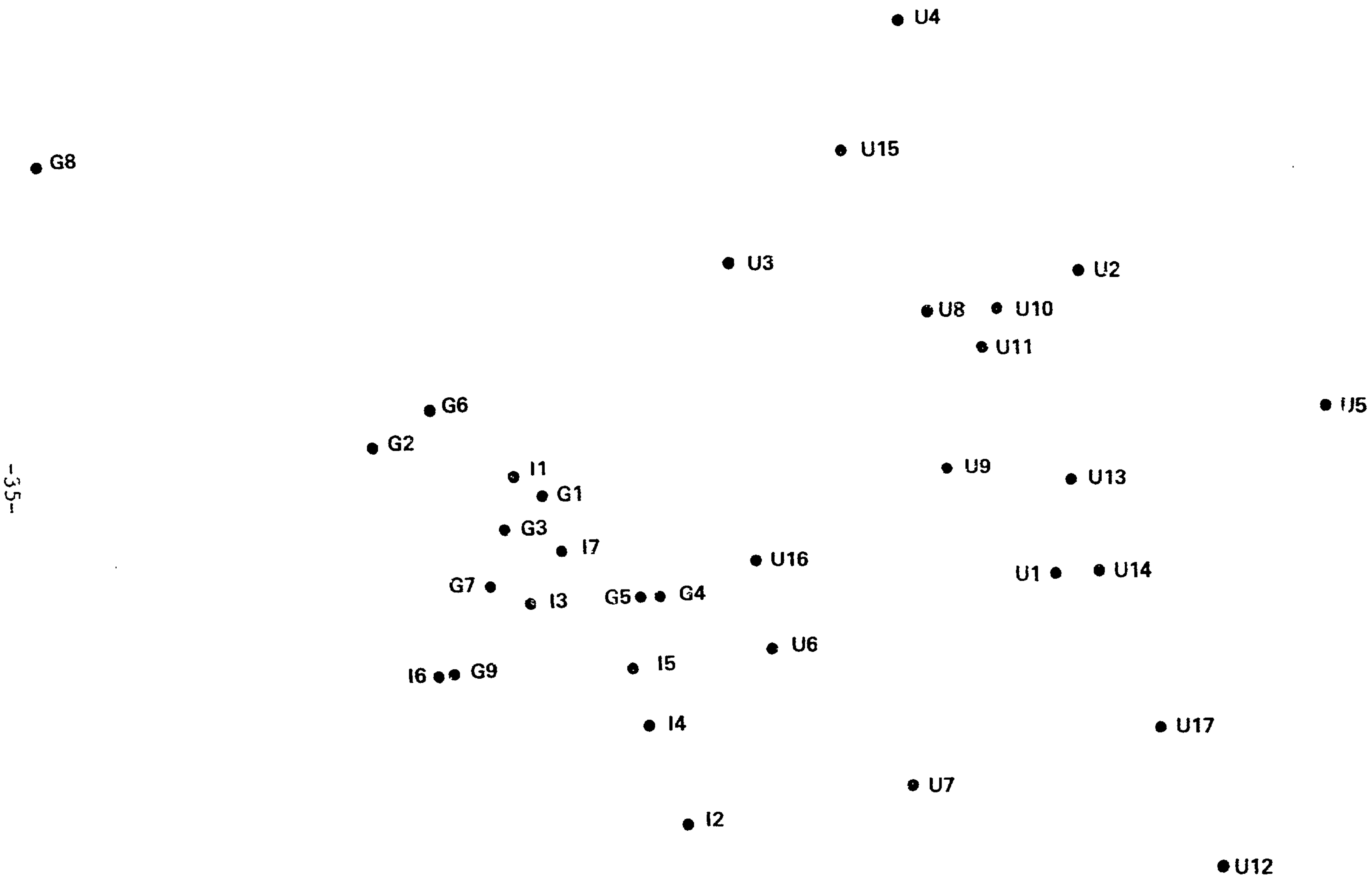

Figure 6. Two-dimensional person space. See text for explanation.

Comparison of techniques. Each of the three techniques just discussed provides information which orders UPs in relation to the experts. The techniques point to a different "best" student. However, inspection of Table 4 also suggests that there is substantial agreement among the three techniques in their ordering of the UPs. All agree, for example, that UP3 is superior to UP4. Spearman correlations were performed on the ranks of the UPs in order to more rigorously compare the techniques. The intercorrelation matrix appears in Table 5. All correlations were reliable. As can be seen, there is a good deal of agreement among the techniques. Pattern classification, INDSCAL, and the person space developed here supply converging validation that may be useful in the selection of students. Further, the techniques supply additional, different details about the populations that may further facilitate selection or aid in training.

Training

In addition to implications for personnel and selection, an understanding of the cognitive structures of experts and novices should have implications for training. Although the underlying structures of some students are closer than are others to those of experts, it should be possible to facilitate acquisition of the expert structure for the students in general. Accomplishing this requires that it first be determined which concepts are not well understood by the UPs relative to expert fighter pilots. This requires that the information critical to expertise be determined. This is not a trivial problem because any individual expert will tend to have associations that, while perhaps useful, are not necessary for expertise. Thus, the problem is to determine which associations in an expert knowledge base are necessary or essential for expertise and which are not. Again the scaling procedures considered in this report have proven of great utility.

Critical information is defined as those components of the cognitive structures that tend to be present in all experts. Any information contained in the knowledge structure of one group of experts, but not another, cannot be a prerequisite component of expertise. After the information critical to expertise is established, one can compare the UPs to the experts, and isolate the concepts that have been mastered by the students. In addition, those concepts which are the most disparate from those of the experts can be determined and thus provide some information concerning which deficits should be addressed first in any pedagogical intervention.

Table 5. Spearman Intercorrelations Matrix of Three Individual Difference Measures

	Person Space PS	INDSCAL IN	Pattern Classification PC
PS	1.00	.63	.92
IN	.63	1.00	.74
PC	.92	.74	1.00

Characteristics of Expertise. The split-plane data of the two groups of experts, IPs and GPs, were considered first. Defining expertise based on these two groups has several advantages. Most importantly, comparing two relatively different groups that are both expert reduces the likelihood that idiosyncratic components of the cognitive structures will manifest themselves as critical components of expertise. As we showed earlier, IPs could be distinguished from GPs; although, of course, the distinction was not as great as between UPs and either of our expert groups. The information common to IPs and GPs should be the minimal structure necessary for expertise.

The GWN analyses were used to determine those features of networks that tended to be characteristic of expertise. The networks for IPs and for GPs considered earlier were compared and any common links were extracted. The resulting network consisted of one major network, three isolated concept pairs, and five isolated single concepts. Thus, IPs and GPs agreed on a way to interconnect 19 of the split-plane concepts and agreed on an additional three pairwise associations; IPs and GPs did not agree on any particular link for five concepts. Then the isolated pairs were linked to the main network by allowing either member of the pair to connect to a particular concept in the main network. For example, IPs and GPs agreed that AIRSPEED is related to the pair ACCELERATION-EXTENSION; however, IPs had AIRSPEED and ACCELERATION linked, whereas GPs linked AIRSPEED with EXTENSION. A similar procedure linked the isolated single concepts to the main network. The resulting structure of "expertise" is shown in Figure 7.

Novices and experts. Next, the network for UPs was compared with the expert structure. A bold line in Figure 7 represents a link present in the UP network. As can be seen, a number of critical links are also held by UPs. These links center around the concept of airspeed and, to a lesser extent, the concept of guns. Several links are not present in the UP network. For example, the links near the concept lag roll are almost totally absent from the UP network.

In order to quantify UPs' understanding of the concepts, each of the 30 split-plane concepts was considered individually. For each concept, students might differ from experts in two general ways. Students may not have some of the critical associations that experts have (as illustrated in Figure 7). Alternatively, students may have associations between concepts that neither group of experts has (e.g., the network of UPs has a link between HIGH YO YO and LOW YO YO, although for any expert these concepts are relatively unrelated). These two general dimensions can combine to produce four different types of concepts: (a) a concept can be well defined in that the critical links are present and the student does not see many spurious additional relations, (b) a concept can be overdefined in that the critical links are present, but the student also has a number of inappropriate links, (c) a concept can be underdefined in that the critical links are absent but so are idiosyncratic links,

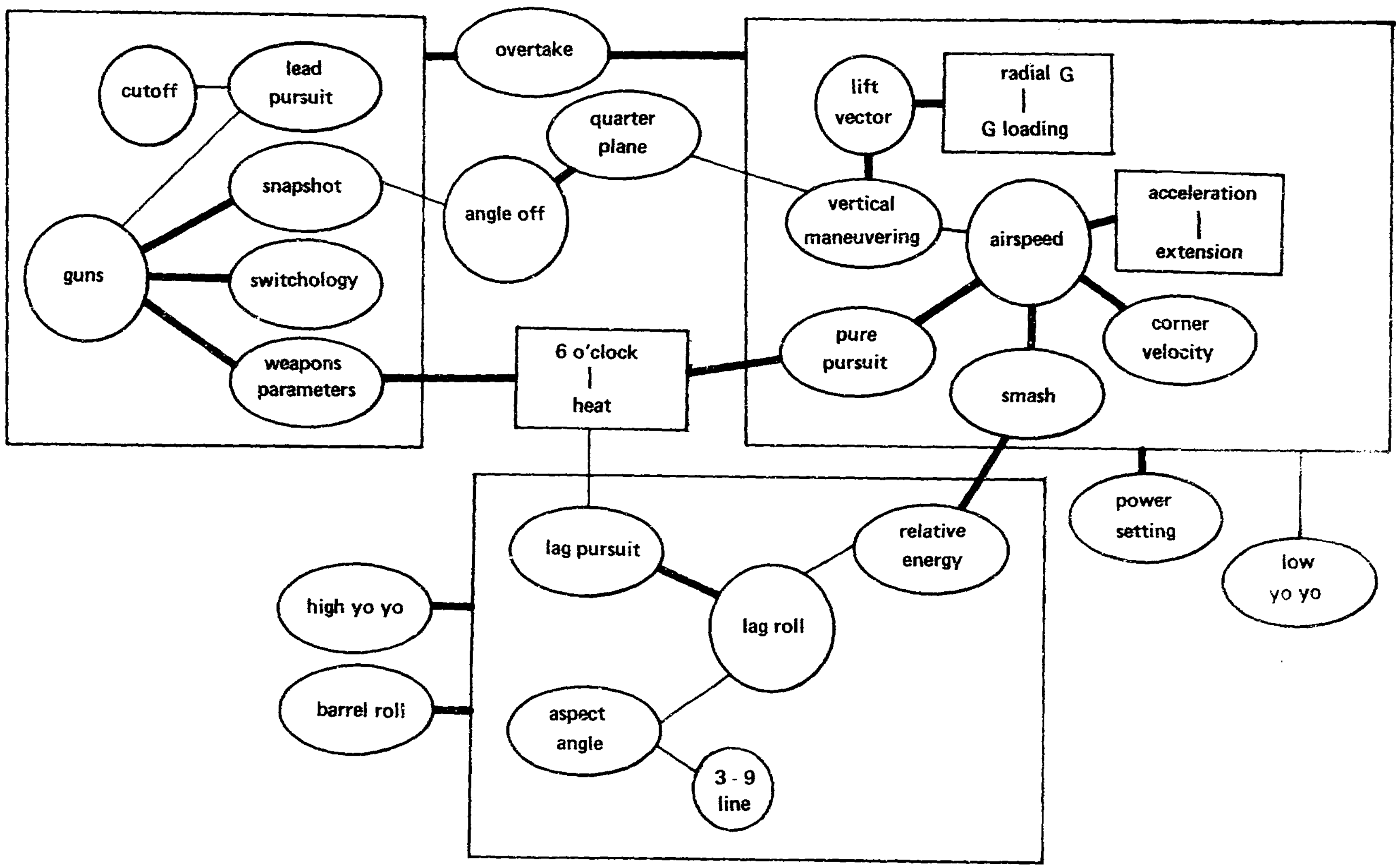

Figure 7. Network overlap for instructor and guard pilots. Bold links are also held by undergraduates.

(d) a concept can be misdefined in that the critical links are absent <u>and</u> the student has many idiosyncratic, nonexpert connections.

The following values were computed for each concept: (a) the proportion of critical links found in the UP network and (b) the proportion of the UP links that were found only in the UP networks (i.e., the "extra" links that occurred in neither the IP nor the GP network). A median split along each variable led to the classification of concepts appearing in Table 6. The well defined concepts tend to be those involved in flying the aircraft, with some consideration of the aircraft's tactical functions, but no terms showing an understanding of split-plane maneuvers or an understanding of air-to-air combat scenarios. Interestingly, these well defined concepts probably evoke a reasonably accurate meaning from a reader naive to tactical flight procedures.

To supply some converging validation of this division, the MDS distances were examined on a concept-by-concept basis. As with the network analysis, an expert space was computed first, based on the mean MDS distances of the IPs and GPs. Then for each concept, for both the experts and the UPs, a vector was created for the distances from the target concept to the 29 other concepts. Then, the expert distance vector was correlated with the UP distance vector for each of the 30 concepts. High postive correlations suggest that the UPs have acquired a <u>global</u> understanding of the concept. Negative correlations, or low positive correlations, suggest that the UPs have not acquired full understanding of the concept (or at least that this understanding is different from that of the experts). In general, the well defined concepts were expected to show high positive correlations relative to the other categories, and the misdefined concepts were expected to show low correlations. Table 6 includes the $\underline{r}$ obtained for each concept and the mean familiarity rating.

As Table 6 suggests, the concepts classified as well defined based on the GWN analyses also tend to show high correlations in multidimensional space. The remaining categories have substantially lower correlations, with misdefined being somewhat lower than either overdefined or underdefined.

Students' judgments of familiarity are also reasonably consistent with the classification in Table 6. Well defined concepts are the most familiar with misdefined concepts the least familiar. However, familiarity does not always agree with the GWN and MDS analyses. For example, some underdefined concepts (CUTOFF and VERTICAL MANEUVERING) and some misdefined concepts (ACCELERATION, 6 O'CLOCK, RELATIVE ENERGY) are judged as very familiar. For each of these "familiar" concepts, it seems the student is not aware of the true scope of the concept. The student seems to have an understanding of the concept in a narrower sense than does the expert. Inspection of Figure 7 reveals that each of these concepts has a critical connection in the expert structure to a concept that the students

Table 6. Thirty Split-Plane Concepts Grouped According to Network Analysis with Correlations between UP and Expert (IPs + GPs) Multidimensional Spaces

	r	Familiarity
Well Defined Concepts		
AIRSPEED	.85	3.00
OVERTAKE	.84	3.00
G-LOADING	.63	2.94
SMASH	.82	2.81
LIFT VECTOR	.74	2.75
CORNER VELOCITY	.75	2.38
GUNS	.76	2.00
SWITCHOLOGY	.46	1.81
Mean	.73	2.59
Underdefined Concepts		
CUTOFF	.20	3.00
VERTICAL MANEUVERING	.63	2.88
ANGLE-OFF	.61	2.00
WEAPONS PARAMETERS	.81	1.75
PURE PURSUIT	.49	1.56
LEAD PURSUIT	.28	1.50
ASPECT ANGLE	.67	1.19
QUARTER PLANE	.50	1.13
Mean	.52	1.88
Over Defined Concepts		
POWER SETTING	.78	3.00
BARREL ROLL	.63	2.94
HI YO YO	.14	1.50
RADIAL G	.19	1.31
Mean	.44	2.19
Misdefined Concepts		
ACCELERATION	.80	2.94
6 O'CLOCK	.50	2.81
RELATIVE ENERGY	.85	2.75
LAG PURSUIT	.22	1.50
LOW YO YO	.22	1.38
3-9 LINE	.59	1.25
SNAPSHOT	.55	1.19
LAG ROLL	.72	1.19
HEAT	.25	1.19
EXTENSION	-.43	1.19
Mean	.43	1.74

have not experienced. For example, whereas students are familiar with ACCELERATION and, in fact, have a global understanding of it, they are missing the critical connection with EXTENSION, a concept with which they have had little or no experience. Thus, UPs have an understanding of ACCELERATION in the same sense that a psychology undergraduate, who knows nothing of ANOVA, might have an understanding of Student's _t_.

If UPs do have much of the cognitive structure of the experts for the well defined concepts, and little of the expert structure for the misdefined concepts, then classification based on these subsets of concepts should reflect the difference in understanding. Classification of UPs and experts based only on the well defined concepts should be relatively poor because there would be little information in these concepts that would allow classification. On the other hand, classification based on the misdefined concepts should be relatively successful because the UPs differ a great deal from the experts.

Pattern classification of UPs and experts was performed using each of the four subsets of concepts. Minimum distance classification results appear in Table 7. Again this classification is particularly easy, even when using only the minimum distance classification (i.e., the intitial step in classification) and only a subset of the concepts. However, consistent with expectations, there were differences at this high level of success: the well defined concepts yielded poorest classification and misdefined concepts yielded the best classification. In fact, with only the subset of misdefined concepts the classification was perfect. In addition to supporting Table 6, these results have the practical advantage of allowing classification with only a small portion of the total data set, thus reducing the amount of data that needs to be collected and the run time of the algorithm.

Tables 6 and 7 present some encouraging results: GWN and MDS tend to converge on particular concepts that seem to have been mastered by the UPs. In addition, when the two techniques are considered together, they point out not only the concepts that are not very well understood, but also the kind of misunderstanding present (i.e.,missing critical connections or having too many connections). While some of these "not understood" concepts are to be expected, simply because of the students' lack of experience with them (i.e., low familiarity), a number of concepts which are included in flight training and are familiar to the students are not well defined according to GWN, according to MDS, or according to both. It is for these concepts that classroom intervention may be particularly promising.

Table 7. Results of Pattern Classification Based on Subsets of Concepts from the Split-Plane Scenario (Classification Utilized Distance Vectors from the Three-Dimensional MDS Solution)

Concept subgroup			
Well defined	Overdefined	Underdefined	Misdefined
82%	85%	85%	100%

Summary

Possible applications of the scaling procedures to selection and training have been suggested. Information about conceptual structure might be profitably applied to aid in decisions about assignments to fighter aircraft. Three techniques developed to assess individual differences showed considerable agreement in their suggestions of which undergraduate pilots should be directed to lead-in-fighter training. In addition, the attempts here to apply information about conceptual structure to training suggested particular points in the UPs' understanding that could benefit from intervention. The particular weak points in the knowledge structure was suggested by GWN and MDS, sometimes in the face of the students' self-perception of their familiarity with the concepts. The subset of misdefined concepts contained enough information about the pilots to classify them perfectly.

The success of the methods at discriminating among pilots of varying expertise based on measures of conceptual structure suggests that scaling methods may provide some empirical techniques for measuring the structure of expertise. These techniques should have application in training and selection as well as in artificial intelligence systems that attempt to represent knowledge structures. Ways of using some of these structures in database systems are being investigated.

References

Adelson, B. Problem solving and the development of abstract categories in programming languages. Memory and Cognition, 1981, 9, 422-433.

Carroll, J. D., & Chang, J. J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. Psychometrica, 1970, 35, 283-319.

Chase, W. G., & Simon, H. A. Perception in chess. Cognitive Psychology, 1973, 4, 55-81.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. Categorization and representation of physics problems by experts and novices. Cognitive Science, 1981, 5, 121-152.

Christofides, N. Graph theory: An algorithmic approach. New York: Academic Press, 1975.

Collins, A. M., & Loftus, E. F. A spreading activation theory of semantic processing. Psychological Review, 1975, 82, 407-428.

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 240-247.

Cooke, N. M. Memory structures of expert and novice computer programmers: Recall order vs. similarity ratings. Unpublished Masters Thesis, New Mexico State University, 1983.

Durso, F. T., Schvaneveldt, R. W., & Goldsmith, T. E. Empirical networks of natural categories. Paper presented at the meeting of the Southwestern Psychological Association, San Antonio, April 1983.

Engle, R. W., & Bukstel, L. Memory processes among bridge players of differing expertise. American Journal of Psychology, 1978, 91, 673-689.

Fillenbaum, S., & Rapaport, A. Structures in the subjective lexicon. New York: Academic Press, 1971.

Hutchinson, J. W. Network representations of psychological relations. Unpublished Doctoral dissertation, Stanford University, 1981

Isaac, P. D., & Poor, D. S. On the determination of appropriate dimensionality in data with error. Psychometrica, 1974, 39, 91-109.

Kruskal, J. B. Multidimensional scaling and other methods for discovering structure. In Enslein, Ralston, and Wilf (Eds.), _Statistical methods for digital computers_. New York: Wiley, 1977.

McKeithen, K. B., Reitman, J. S., Rueter, H. H., & Hirtle, S. C. Knowledge organization and skill differences in computer programmers. _Cognitive Psychology_, 1981, _13_, 307-325.

Meyer, R. P., Laveson, J. I., Pape, G. L., & Edwards, B. J. _Development and application of a task taxonomy for tactical flying_. AFHRL-TR-78-42(III), AD-A061 478. Williams AFB, AZ: Flying Training Division, Air Force Human Resources Laboratory, September 1978.

Nilsson, N. _Learning machines: Foundations of trainable pattern-classification systems_. New York: McGraw-Hill, 1965.

Quillian, M. R. The teachable language comprehender. _Communications of the Association for Computing Machinery_, 1969, _12_, 459-475.

Reitman, J. S. Skilled perception in Go: Deducing memory structures from interresponse times. _Cognitive Psychology_, 1976, _8_, 336-356.

Rips, L. J., Shoben, E. J., & Smith, E. E. Semantic distance and the verification of semantic relations. _Journal of Verbal Learning and Verbal Behavior_, 1973, _12_, 1-20.

Schvaneveldt, R. W., & Durso, F. T. _Generalized semantic networks_. Paper presented at the meeting of the Psychonomic Society, Philadelphia, November 1981.

Shepard, R. N. Analysis of proximities as a technique for the study of information processing in man. _Human Factors_, 1963, _5_, 33-48.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A featural model of semantic decisions. _Psychological Review_, 1974, _81_, 214-241.

Tversky, A. Features of similarity. _Psychological Review_, 1977, _84_, 327-351.